Claims

1. A method of anomaly detection characterised in that it incorporates the steps of:-

   a) developing a rule set of at least one anomaly characterisation rule from a training data set and any available relevant background knowledge using at least first order logic, a rule covering a proportion of positive anomaly examples of data in the training data set, and

   b) applying the rule set to test data for anomaly detection therein.

2. An automated method of anomaly detection characterised in that it comprises using computer apparatus to execute the steps of:-

   a) developing a rule set of at least one anomaly characterisation rule from a training data set and any available relevant background knowledge using at least first order logic, a rule covering a proportion of positive anomaly examples of data in the training data set, and

   b) applying the rule set to test data for anomaly detection therein.

3. A method according to Claim 2 characterised in that the positive anomaly examples are associated with fraud or software vulnerabilities.

4. A method according to Claim 2 characterised in that it includes developing the rule set using Higher-Order logic.

5. A method according to Claim 4 characterised in that it includes developing the rule set by:

   a) forming an alphabet having selector functions allowing properties of the training data set to be extracted, together with at least one of the following: additional concepts, background knowledge constant values and logical AND and OR functions,

   b) forming current rules from combinations of items in the alphabet such that type consistency and variable consistency is preserved,

   c) evaluating the current rules for adequacy of classification of the training data

set,

d) if no current rule adequately classifies the training data set, generating new rules by applying at least one genetic operator to the current rules, a genetic operator having one of the following functions: i) combining two rules to form a new rule, ii) modifying a single rule by deleting one of its conditions or adding a new condition to it, or iii) changing one of a rule's constant values for another of an appropriate type, and

e) designating the new rules as the current rules and iterating steps c) onwards until a current rule adequately classifies the training data set or a predetermined number of iterations is reached.

6. A method according to Claim 2 characterised in that data samples in the training data set have characters indicating whether or not they are associated with anomalies.

7. A method according to Claim 6 characterised in that it is a method of detecting telecommunications or retail fraud from anomalous data.

8. A method according to Claim 7 characterised in that it employs inductive logic programming to develop the rule set.

9. A method according to Claim 8 characterised in that the at least one anomaly characterisation rule has a form that an anomaly is detected or otherwise by application of the rule according to whether or not a condition set of at least one condition associated with the rule is fulfilled.

10. A method according to Claim 9 characterised in that the at least one anomaly characterisation rule is developed by refining a most general rule by at least one of:

a) addition of a new condition to the condition set; and

b) unification of different variables to become constants or structured terms.

11. A method according to Claim 10 characterised in that a variable in the at least one anomaly characterisation rule which is defined as being in constant mode and is numerical is at least partly evaluated by providing a range of values for the variable, estimating an accuracy for each value and selecting a value having optimum

accuracy.

12. A method according to Claim 11 characterised in that the range of values is a first range with values which are relatively widely spaced, a single optimum accuracy value is obtained for the variable, and the method includes selecting a second and relatively narrowly spaced range of values in the optimum accuracy value's vicinity, estimating an accuracy for each value in the second range and selecting a value in the second range having optimum accuracy.

13. A method according to Claim 12 characterised in that it includes filtering to remove rule duplicates and rule equivalents, i.e. any rule having like but differently ordered conditions compared to another rule, and any rule which has conditions which are symmetric compared to those of another rule.

14. A method according to Claim 13 characterised in that it includes filtering to remove unnecessary 'less than or equal to' ("lteq") conditions.

15. A method according to Claim 14 characterised in that the unnecessary "lteq" conditions are associated with at least one of ends of intervals, multiple lteq predicates and equality condition and lteq duplication.

16. A method according to Claim 8 characterised in that it includes implementing an encoding length restriction to avoid overfitting noisy data by rejecting a rule refinement if the refinement encoding cost in number of bits exceeds a cost of encoding the positive examples covered by the refinement.

17. A method according to Claim 8 characterised in that it includes stopping construction of a rule if at least one of three stopping criteria is fulfilled as follows:

a) the number of conditions in any rule in a beam of rules being processed is greater than or equal to a prearranged maximum rule length,

b) no negative examples are covered by a most significant rule, which is a rule that:

i) is present in a beam currently being or having been processed,

ii) is significant,

iii) has obtained a highest likelihood ratio statistic value found so far, and

iv) has obtained an accuracy value greater than a most general rule accuracy value, and

c) no refinements were produced which were eligible to enter the beam currently being processed in a most recent refinement processing step (32).

18. A method according to Claim 17 characterised in that it includes adding the most significant rule to a list of derived rules and removing positive examples covered by the most significant rule from the training data set.

19. A method according to Claim 8 characterised in that it includes:

a) selecting rules which have not met rule construction stopping criteria,

b) selecting a subset of refinements of the selected rules associated with accuracy estimate scores higher than those of other refinements of the selected rules, and

c) iterating a rule refinement, filtering and evaluation procedure (32 to 38) to identify any refined rule usable to test data.

20. Computer apparatus for anomaly detection characterised in that it is programmed to execute the steps of:-

a) developing a rule set of at least one anomaly characterisation rule from a training data set and any available relevant background knowledge using at least first order logic, a rule covering a proportion of positive anomaly examples of data in the training data set, and

b) applying the rule set to test data for anomaly detection therein.

21. Computer apparatus according to Claim 20 characterised in that the positive anomaly examples are associated with fraud or software vulnerabilities.

22. Computer apparatus according to Claim 20 characterised in that it is programmed to develop the rule set using Higher-Order logic.

23. Computer apparatus according to Claim 22 characterised in that it includes developing

the rule set by:

a) forming an alphabet having selector functions allowing properties of the training data set to be extracted, together with at least one of the following: additional concepts, background knowledge constant values and logical AND and OR functions,

b) forming current rules from combinations of items in the alphabet such that type consistency and variable consistency is preserved,

c) evaluating the current rules for adequacy of classification of the training data set,

d) if no current rule adequately classifies the training data set, generating new rules by applying at least one genetic operator to the current rules, a genetic operator having one of the following functions: i) combining two rules to form a new rule, ii) modifying a single rule by deleting one of its conditions or adding a new condition to it, or iii) changing one of a rule's constant values for another of an appropriate type, and

e) designating the new rules as the current rules and iterating steps c) onwards until a current rule adequately classifies the training data set or a predetermined number of iterations is reached.

24. Computer apparatus according to Claim 20 characterised in that data samples in the training data set have characters indicating whether or not they are associated with anomalies.

25. Computer apparatus according to Claim 20 characterised in that the at least one anomaly characterisation rule has a form that an anomaly is detected or otherwise by application of such rule according to whether or not a condition set of at least one condition associated with that rule is fulfilled.

26. Computer apparatus according to Claim 20 characterised in that the at least one anomaly characterisation rule is developed by refining a most general rule by at least one of:

a) addition of a new condition to the condition set; and

b) unification of different variables to become constants or structured terms.

27. Computer apparatus according to Claim 26 characterised in that a variable in the at least one anomaly characterisation rule which is defined as being in constant mode and is numerical is at least partly evaluated by providing a range of values for the variable, estimating an accuracy for each value and selecting a value having optimum accuracy.

28. Computer apparatus according to Claim 25 characterised in that it is programmed to filter out at least one of rule duplicates, rule equivalents and unnecessary 'less than or equal to' ("lteq") conditions.

29. Computer apparatus according to Claim 25 characterised in that it is programmed to stop construction of a rule if at least one of three stopping criteria is fulfilled as follows:

d) the number of conditions in any rule in a beam of rules being processed is greater than or equal to a prearranged maximum rule length,

e) no negative examples are covered by a most significant rule, which is a rule that:

i) is present in a beam currently being or having been processed,

ii) is significant,

iii) has obtained a highest likelihood ratio statistic value found so far, and

iv) has obtained an accuracy value greater than a most general rule accuracy value, and

f) no refinements were produced which were eligible to enter the beam currently being processed in a most recent refinement processing step.

30. Computer software for use in anomaly detection characterised in that it incorporates instructions for controlling computer apparatus to execute the steps of:-

a) developing a rule set of at least one anomaly characterisation rule from a training data set and any available relevant background knowledge using at least first order logic, a rule covering a proportion of positive anomaly examples of data in the training data set, and

b) applying the rule set to test data for anomaly detection therein.

31. Computer software according to Claim 30 characterised in that the positive anomaly

examples are associated with fraud or software vulnerabilities.

32. Computer software according to Claim 30 characterised in that it incorporates instructions for controlling computer apparatus to develop the rule set using Higher-Order logic.

33. Computer software according to Claim 32 characterised in that it incorporates instructions for controlling computer apparatus to develop the rule set by:

a) forming an alphabet having selector functions allowing properties of the training data set to be extracted, together with at least one of the following: additional concepts, background knowledge constant values and logical AND and OR functions,

b) forming current rules from combinations of items in the alphabet such that type consistency and variable consistency is preserved,

c) evaluating the current rules for adequacy of classification of the training data set,

d) if no current rule adequately classifies the training data set, generating new rules by applying at least one genetic operator to the current rules, a genetic operator having one of the following functions: i) combining two rules to form a new rule, ii) modifying a single rule by deleting one of its conditions or adding a new condition to it, or iii) changing one of a rule's constant values for another of an appropriate type, and

e) designating the new rules as the current rules and iterating steps c) onwards until a current rule adequately classifies the training data set or a predetermined number of iterations is reached.

34. Computer software according to Claim 30 characterised in that data samples in the training data set have characters indicating whether or not they are associated with anomalies.

35. Computer software according to Claim 30 characterised in that the at least one anomaly characterisation rule has a form that an anomaly is detected or otherwise by application of such rule according to whether or not a condition set of at least one

condition associated with that rule is fulfilled.

36. Computer software according to Claim 30 characterised in that it incorporates instructions for controlling computer apparatus to develop the at least one anomaly characterisation rule by refining a most general rule by at least one of:

a) addition of a new condition to the condition set; and

b) unification of different variables to become constants or structured terms.

37. Computer software according to Claim 36 characterised in that it incorporates instructions for controlling computer apparatus to at least partly evaluate a variable in the at least one anomaly characterisation rule which is defined as being in constant mode and is numerical by providing a range of values for the variable, estimating an accuracy for each value and selecting a value having optimum accuracy.

38. Computer software according to Claim 35 characterised in that it incorporates instructions for controlling computer apparatus to filter out at least one of rule duplicates, rule equivalents and unnecessary 'less than or equal to' ("lteq") conditions.

39. Computer software according to Claim 35 characterised in that it incorporates instructions for controlling computer apparatus to stop construction of a rule if at least one of three stopping criteria is fulfilled as follows:

g) the number of conditions in any rule in a beam of rules being processed is greater than or equal to a prearranged maximum rule length,

h) no negative examples are covered by a most significant rule, which is a rule that:

i) is present in a beam currently being or having been processed,

ii) is significant,

iii) has obtained a highest likelihood ratio statistic value found so far, and

iv) has obtained an accuracy value greater than a most general rule accuracy value, and

i) no refinements were produced which were eligible to enter the beam currently being processed in a most recent refinement processing step.